Memory Map between External Memory and On-Chip Memory

# 概要

配置寄存器数据、原始图像数据、权重数据、中间层输出等数据需要在外部存储空间和OCM(On-Chip Memory)之间通过DMA进行搬移。分5种场景：

1. Scenario1：Configuration registers。单向，从外部搬移到OCM；
2. Scenario2：Weights。单向，从外部搬移到OCM。在写入OCM之前需要解压缩；
3. Scenario3：Input Images。单向，从外部搬移到OCM。由于原始图像CHANNEL数一般为3，写入OCM之前要经过reshape处理，使CHANNEL能够被16整除；
4. Scenario4：Feature maps。双向，从外部搬移到OCM，或OCM到外部。中间层输出在OCM存不下时需要存到外部存储空间再分批读进OCM；
5. Scenario5：中间层数据+残差。单向，从外部搬移到OCM，写入OCM之前作残差运算。

本文对以上DMA地址映射（MEMORY MAP）关系作说明。

# 变量名称

为方便描述，定义如下变量。本文中以BITS=128来说明。

表格 1 变量名称

|  |  |  |
| --- | --- | --- |
| 序号 | 变量名称 | 含义 |
|  | BITS | OCM word bits width |
|  | W | Input feature map Width |
|  | H | Input feature map Height |
|  | C | Input feature map Channel |
|  | M | Output feature map Channel |
|  | R | Kernel Width |
|  | S | Kernel Height |
|  | src\_sa | source start address |
|  | dst\_sa | destination start address |

# Memory Map Scenarios

以下说明基于BITS=128。

## MM of configuration registers（Scenario 1）

假设有N个128-bit字，标记为Reg0，Reg1，…，Reg N-1。Scenario1为顺序地址映射，如图 1所示。

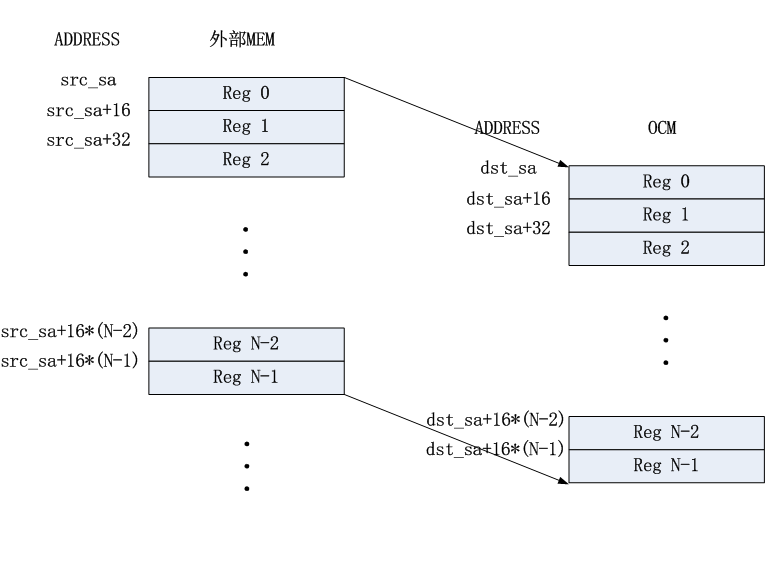


图 1 MM of Scenario 1

## MM of Weights data（Scenario 2）

假设C=32，M=16，R=2，S=2，BITS=128。如图 2，共有128个128-bit字，记为Weight 0，Weight 1，…，Weight 127。

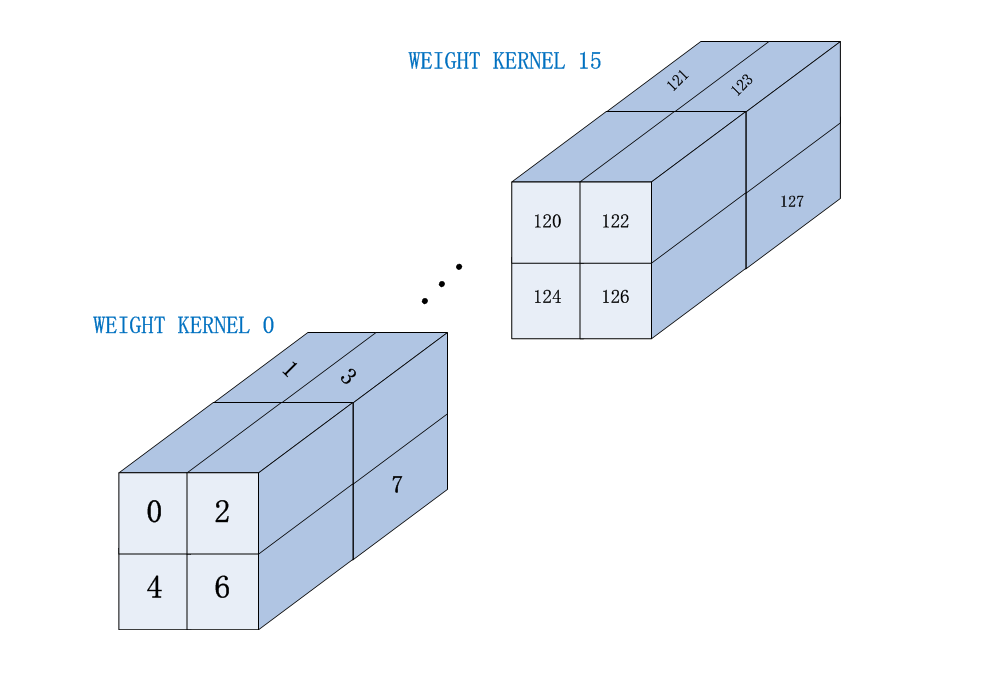


图 2 Weight kernel in 3-D vision

Weights Data在外部MEM与OCM之间为顺序地址映射，如图 3所示。

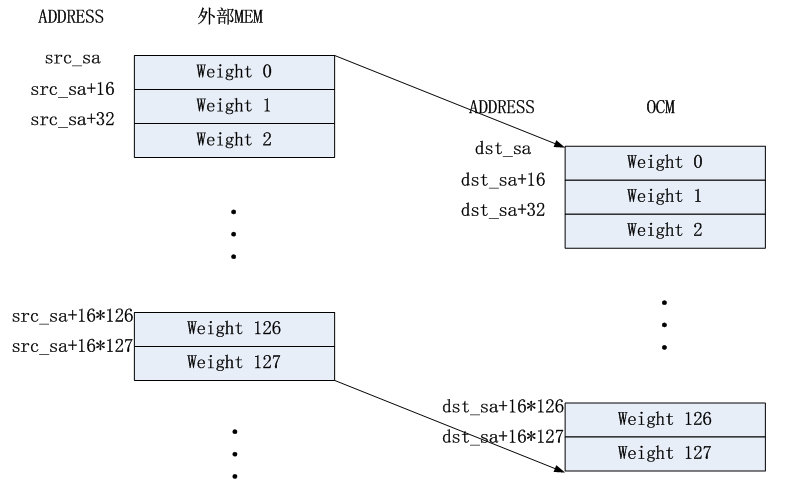


图 3 MM of scenario 2

## MM of Image Data + RESHAPE（Scenario 3）

如图 4所示，原始图像尺寸为(W, H, C)=(8, 8, 3)，共192个像素，记为Pixel 0，Pixel 1，…，Pixel 191。

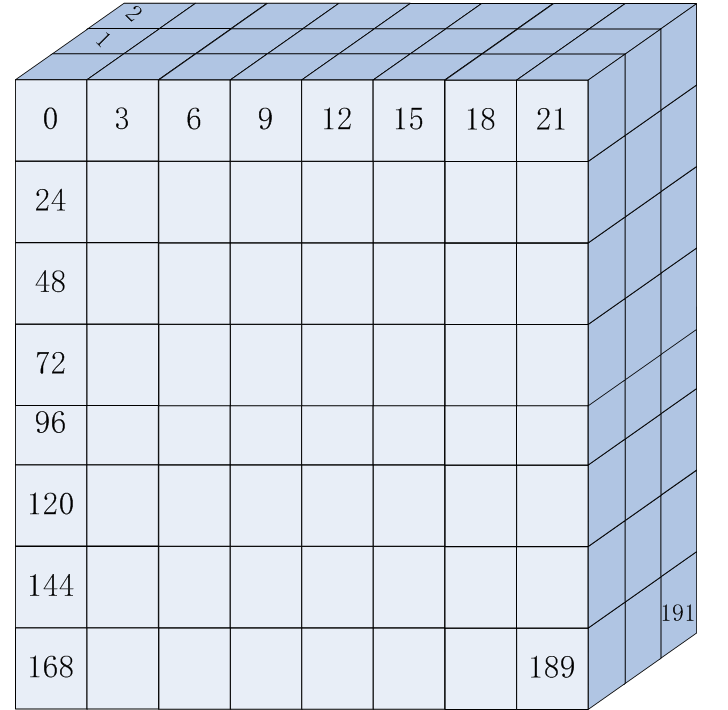


图 4 Original Image. (W, H, C) = (8, 8, 3)

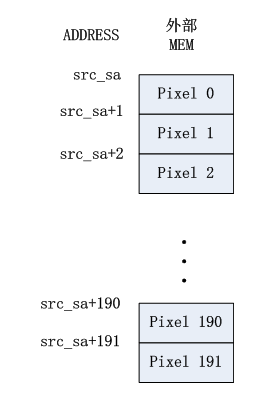


图 5 MM of image data in external memory

假设R=5，S=5，Stride=1，PAD为零，则reshape处理之后如图 6所示，共32个128-bit字，记为0，1，…，31。在OCM中的存储顺序如图 7。

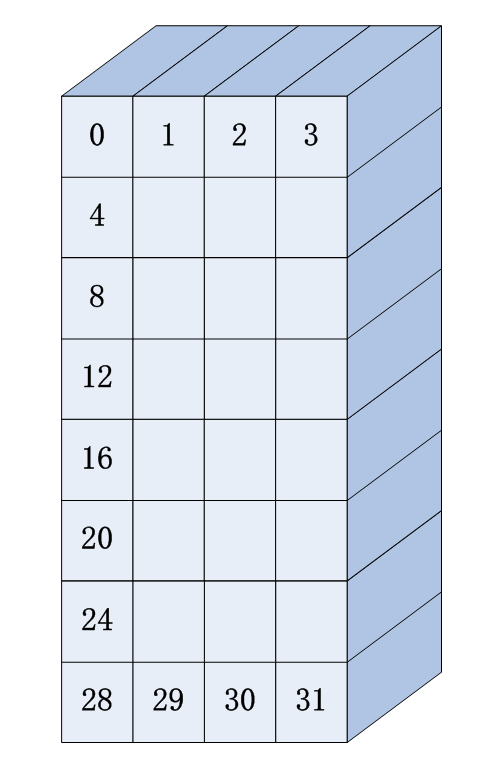


图 6 Original Image after reshape

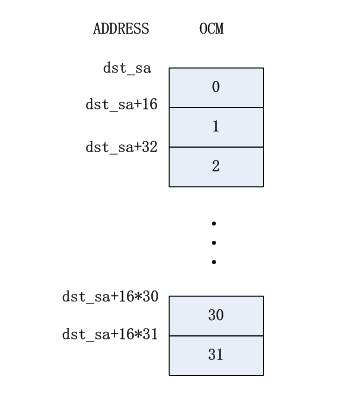


图 7 MM of reshaped image in OCM

假设R=7，S=7，Stride=1，PAD为零，则reshape处理之后如图 8所示，共32个128-bit字，记为0，1，…，31。在OCM中的存储顺序同图 7。

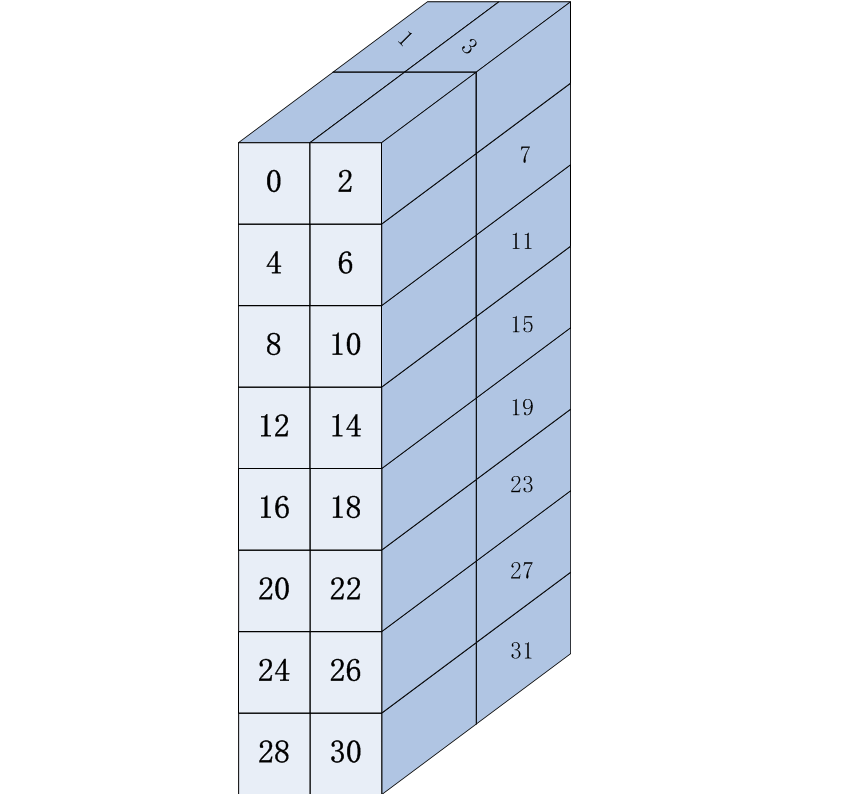


图 8 Original Image after another reshape

## MM of Feature Maps Data（Scenario 4）

假设Feature map (W, H, C)=(2, 2, 32)，如图 9，共8个128-bit字，记为Word 0，Word 1，…，Word 7。外部MEM与OCM之间的地址映射关系如图 10。

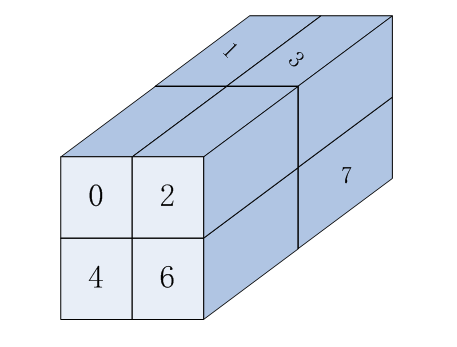


图 9 Feature map (W, H, C) = (2, 2, 32)

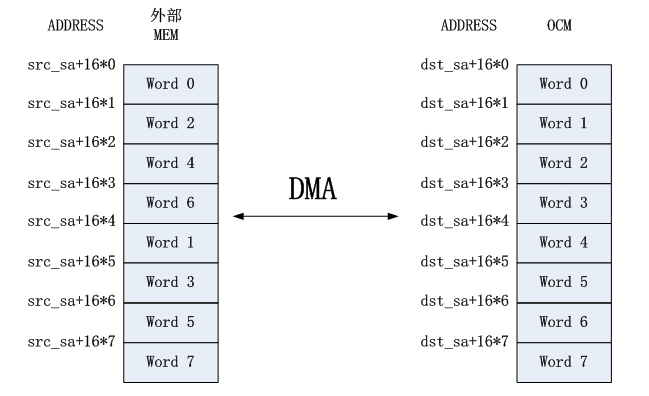


图 10 MM of Feature Map data between external and OCM

MM of Feature Maps Data（Scenario 4）

## MM of Feature Maps data + SHORTCUT

如图 11，2个尺寸为(W, H, C) = (2, 2, 32)的Feature map拼接到一起，各自包含8个128-bit字，记为Word0，Word1，…，Word7 和Word0’，Word1’, …，Word7’，对应Word作残差之后的到如图 9所示的数据，缩减为8个128-bit字，记为New Word0，New Word1，…，New Word7。地址映射关系如所图 12示。

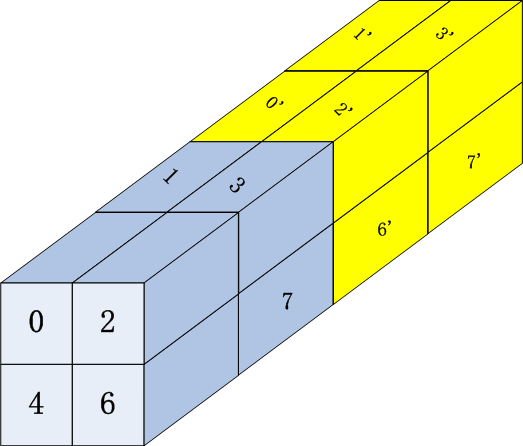


图 11 2 Feature maps concatenated, each with (W, H, C) = (2, 2, 32)

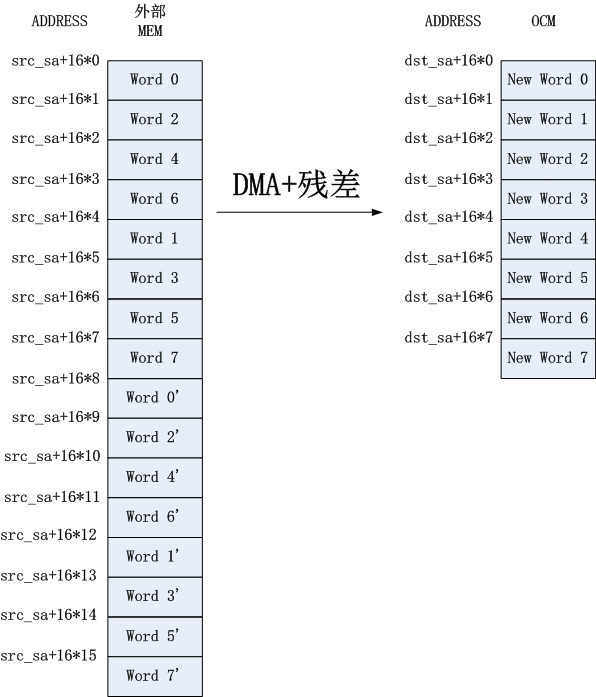


图 12 MM of Feature Map after shortcut